

CLAIMS**We claim:**

1. A system for performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through
5 an ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith, the system comprising:
 - a first domain name server deployed on an Internet backbone (DNS-B); and
 - a plurality of load balancing domain name servers (DNS-LBs) deployed in
10 close physical proximity to the clients, the DNS-LBs having stored therein IP address information of the multiple globally-dispersed servers to be load balanced, the DNS-LBs each sending mapping information to the DNS-B relating the DNS-LB's IP address to an IP address of the DNS-ISP to which it is in close physical proximity, the DNS-LBs determining performance characteristics of each of the multiple globally-
15 dispersed servers.
2. The system of claim 1, wherein the DNS-B stores the mapping information for the plurality of DNS-LBs to forward IP address queries to one of the DNS-LBs closest to the DNS-ISP from which the IP address query originated, and
20 wherein the DNS-LB closest to the DNS-ISP returns the IP address to the DNS-ISP of the server having the best performance characteristics.
3. The system of claim 1, wherein the DNS-B stores the mapping information for the plurality of DNS-LBs to forward IP address queries to one of the DNS-LBs closest

to the DNS-ISP from which the IP address query originated, and wherein the DNS-LB closest to the DNS-ISP returns the IP address of the DNS-LB to the DNS-ISP.

4. The system of claim 1, wherein the DNS-B provides its IP address information to the DNS-A to enable the DNS-A to forward IP address queries to the
5 DNS-B.

5. The system of claim 4, wherein the DNS-B receives IP address information from the DNS-A for the servers to be load balanced.

10 6. The system of claim 1, wherein the DNS-LB is a client of the DNS-ISP.

7. The system of claim 1, further comprising a DNS-B deployed on each Internet backbone, and wherein each DNS-B contains the mapping information for all
15 of the DNS-LBs stored therein.

8. The system of claim 1, wherein the DNS-LB transmits updated mapping information upon a change of an IP address of the DNS-ISP.

20 9. The system of claim 1, wherein each of the DNS-LBs transmit performance information of the servers to the DNS-B, and wherein the DNS-B utilizes the mapping information to determine the proper DNS-LB performance information to utilize to select the IP address of the server having the best
25 originated.

10. A method of performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through an ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith, the method comprising the steps of:

receiving IP address information from the DNS-A for the servers to be load balanced;

providing the IP address information to a plurality of load balancing domain name servers (DNS-LB);

receiving mapping information associating DNS-ISP IP address information to IP address information of a proximately located DNS-LB capable of determining server performance from a location physically proximate to the ISP's point of presence; and

referring address inquiries from a DNS-ISP to a physically proximate DNS-LB in accordance with the mapping information.

11. A computer-readable medium having computer executable-instructions for performing the steps of claim 10.

20

12. A method of performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through an ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith, the method comprising the steps of:

25

obtaining, by a load balancing domain name server (DNS-LB), IP address information for a DNS-ISP located in close physical proximity to the DNS-LB;

providing a mapping of an IP address of the DNS-LB to the IP address information of the DNS-ISP to an external domain name server;

- 5 receiving IP address information for the servers;
monitoring performance of the servers at the received IP addresses; and
providing at least one IP address for a server in response to a name query selected based on the monitoring step.

- 10 13. The method of claim 12, further comprising the steps of:
detecting a change in the DNS-ISP IP address; and
updating the mapping of the IP address of the DNS-LB to the IP address information of the DNS-ISP to the external domain name server.

- 15 14. The method of claim 12, further comprising the steps of
receiving selection criteria for the selection of an IP address;
receiving a name query from the DNS-ISP; and
wherein the step of providing at least one IP address for a server in response to
a name query selected based on the monitoring step further comprises the step of
20 providing at least one IP address for a server in response to a name query selected
based on the monitoring step and on the selection criteria.

15. A computer-readable medium having computer-executable instructions for performing the steps of claim 12.

16. A method of performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through an ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith, the method

5 comprising the steps of:

receiving an IP address of a domain name server to which name queries are to be forwarded for load balancing;

providing IP address information for the servers to the domain name server;
and

10 forwarding name queries from the DNS-ISP to the IP address of the domain name server.

17. A computer-readable medium having computer-executable instructions for performing the steps of claim 16.

15

18. A computer-readable medium having stored thereon a data structure, comprising:

a first data field containing IP address information for an Internet service provider's domain name server (DNS-ISP);

20 a second data field containing IP address information for a load balancing domain name server (DNS-LB); and

wherein the first data field and the second data field are associated to one another so as to provide a mapping from the IP address of the DNS-ISP to the IP address of the DNS-LB.

25

19. The computer-readable medium of claim 17, wherein the first data field comprises a first column of data fields containing a plurality of IP address entries for a plurality of DNS-ISPs, wherein the second data field comprises a second column of data fields containing a plurality of IP address entries for a plurality of DNS-LBs, and wherein an association between each entry in the first column and each entry in the second column forms mapping rows from each DNS-ISP IP address to a DNS-LB IP address.

20. A method of performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through Internet service providers (ISPs) at a point of presence (POP), each ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith containing information regarding the IP addresses of the servers, the method comprising the steps of:

15 deploying a first plurality of load balancing domain name servers (DNS-LBs) in close physical proximity to the ISP POPs;

 deploying a second plurality of second level domain name servers (DNS-Bs) on the Internet backbones and regional provides;

 communicating IP address information for the DNS-Bs to the DNS-As to

20 enable the DNS-As to refer name queries to the DNS-Bs;

 providing, by the DNS-LBs to the DNS-B, mapping information associating an IP address of the DNS-LB to an IP address of the physically proximate DNS-ISP to enable the DNS-B to refer name queries from a DNS-ISP to the physically proximate DNS-LB; and

25 communicating IP address information of the servers to the DNS-LBs;

monitoring, by the DNS-LBs at a location physically proximate to the ISP POP, performance of the servers; and

providing, by the DNS-LB in response to a query from the DNS-ISP, the IP address of a server based on the step of monitoring.

5

21. A method of performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through Internet service providers (ISPs) at a point of presence (POP), each ISP having a load balancing domain name server (DNS-ISP-LB), the servers further having an authoritative domain name server (DNS-A) associated therewith containing information regarding the IP addresses of the servers, the method comprising the steps of:

deploying a first plurality of measurement service agents (MServices) in close physical proximity to the ISP POPs;

15 monitoring, by the MServices at a location physically proximate to the ISP POP, performance of the servers; and

providing, by the DNS-ISP-LB in response to a query from the client, the IP address of a server based on the step of monitoring.

20

22. A method of performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through Internet service providers (ISPs) at a point of presence (POP), each ISP having a load balancing domain name server (DNS-ISP-LB), the servers further having an authoritative domain name server (DNS-A) associated therewith containing

information regarding the IP addresses of the servers, the method comprising the steps of:

deploying a first plurality of measurement service agents (MServices) in close physical proximity to the ISP POPs;

5 monitoring, by the MServices at a location physically proximate to the ISP POP, performance of the servers; and

providing, by the DNS-ISP-LB in response to a query from the client, an IP address of the MService.